Erkenntnis manuscript No.
(will be inserted by the editor)

Horizontal Surgicality and Mechanistic Constitution

Michael Baumgartner · Lorenzo Casini · Beate Krickel

Received: date / Accepted: date

Abstract While ideal (surgical) interventions are acknowledged by many as valuable tools for the analysis of causation, recent discussions have shown that, since there are no ideal interventions on upper-level phenomena that non-reductively supervene on their underlying mechanisms, interventions cannot—contrary to a popular opinion—ground an informative analysis of constitution. This has led some to abandon the project of analyzing constitution in interventionist terms. By contrast, this paper defines the notion of a *horizontally surgical intervention*, and argues that, when combined with some innocuous metaphysical principles about the relation between upper and lower levels of mechanisms, that notion delivers a sufficient condition for constitution. This, in turn, strengthens the case for an interventionist analysis of constitution.

Keywords mechanism; constitution; ideal intervention; fat-handed intervention

1 Introduction

The mechanistic view of explanation, which has gained considerable popularity in the philosophy of the special sciences, holds that the upper- (or macro-) level behavior type Ψ of a system s is explained by the lower- (or micro-) level mechanism (type) *constituting* s's Ψ-ing (Glennan, 1996; Machamer et al., 2000; Craver, 2007). Thus,

Michael Baumgartner
Department of Philosophy, University of Bergen, Postboks 7805, 5020 Bergen, Norway
E-mail: michael.baumgartner@uib.no

Lorenzo Casini
Department of Philosophy, University of Geneva, 2 Rue de Candolle, Geneva, CH-1205 Switzerland
E-mail: lorenzo.casini@unige.ch

Beate Krickel
Department of Philosophy II and Graduiertenkolleg/Research Training Group "Situated Cognition", Ruhr-University Bochum, Universitätsstraße 150, Bochum, D-44801 Germany
E-mail: beate.krickel@rub.de

constitution is the key dependence relation in mechanistic explanations. Even though constitution is commonly understood as a non-causal dependence relation (Craver and Bechtel, 2007), the best known theory of constitution, *viz.* Craver's (2007) mutual manipulability theory (MM), defines it by drawing heavily on conceptual resources that have proven valuable in analyzing causation, more precisely on Woodward's (2003) notion of an *ideal intervention*. In short, an ideal intervention surgically fixes the value of exactly one target variable such that all of its other effects, if any, in a scrutinized system are mediated by this target (Woodward, 2003, p. 98; Craver, 2007, p. 154). Against that background, MM stipulates that the behavior type Φ of a spatiotemporal part x of s constitutes (or, synonymously, is constitutively relevant to)[1] s's Ψ-ing iff there exists a possible ideal intervention on x's Φ-ing that changes s's Ψ-ing and a possible ideal intervention on s's Ψ-ing that changes x's Φ-ing (Craver, 2007, p. 159).

Although MM has much intuitive appeal and is still very popular in the literature (see, e.g.: Kaplan, 2012; Irvine, 2013, ch. 6; Zednik, 2015; van Eck and de Jong, 2016), it has recently become clear that it suffers from decisive problems. Most importantly, by definitionally tying constitution to the possible existence of ideal interventions, MM either reduces constitution to causation (Leuridan, 2012), in violation of the widespread view that constitution is a non-causal form of dependence, or entails that cases of constitution cannot possibly exist, for the ideal interventions required by MM are unrealizable in principle (Baumgartner and Gebharter, 2016; Baumgartner and Casini, 2017).[2] In our view, this conclusively establishes that MM is not suited for defining constitution.[3] The question remains, however, whether this finding merely invalidates MM's attempt to spell out constitution in terms of *ideal* interventions or whether it furthermore entails the impossibility of analyzing constitution in terms of *non-ideal* interventions.

Some authors have considered the problems encountered by MM sufficient to abandon the attempt of spelling out constitution in interventionist terms altogether and, consequently, have proposed alternative analyses with alternative methods of constitutional discovery (Harbecke, 2010, 2015; Couch, 2011; Gebharter, 2017). This paper, by contrast, takes a different approach. We explore the possibility of maintaining the original idea of accounting for constitution on interventionist grounds by developing a notion of a non-ideal intervention that, on the one hand, avoids the problems of MM and, on the other, still provides sufficient leverage for inferences to constitution.

As our starting point, we take recent results by Romero (2015), Baumgartner and Gebharter (2016), and Baumgartner and Casini (2017), who have shown that upper

[1] The terms "constitution" and "constitutive relevance" are sometimes used with differing meanings: single constituents are constitutively relevant to the phenomenon while only the whole lower-level mechanism constitutes it. For ease of expression, we use the terms synonymously here, meaning that "Φ constitutes Ψ" is to be understood in terms of "Φ *is a constituent* (among possibly many) of Ψ".

[2] In a recent Stanford Encyclopedia entry, Craver (Craver and Tabery, 2017, sect. 3.2) seems to acknowledge the difficulties in applying Woodward's interventionism to mechanistic systems. However, in an even more recent online presentation, Craver insists that the problems merely concern the formulation of MM in (Craver, 2007) and not the theory's content.

[3] One author (BK) of this paper argues that MM may be saved if the phenomenon is represented by multiple variables standing for different temporal phases of the phenomenon (Krickel, 2018).

and lower levels of mechanisms can only be manipulated via common causes. As a consequence, all interventions that possibly are constitutionally revealing are not surgical but *fat-handed*.[4] Without further restrictions, however, fat-handed interventions systematically underdetermine the inference to constitution (Baumgartner and Casini, 2017). To avoid this underdetermination problem and render unambiguous inferences to constitution possible, we develop an additional constraint that fat-handed interventions on mechanistic systems must satisfy, *viz. horizontal surgicality*. Roughly, fat-handed interventions are horizontally surgical iff they are surgical relative to every level on which they operate, that is, iff they surgically target exactly one variable on every level on which they operate, such that the target constituent(s) and the target phenomenon change their values simultaneously. We shall argue that, under the additional assumption that all changes in upper-level phenomena are necessarily realized by a change in some lower-level constituent or other, a single fat-handed horizontally surgical intervention can conclusively establish that its lower-level target constitutes its upper-level target.

It is then tempting to stipulate that this sufficient condition—*viz.* the existence of a change in the phenomenon under a horizontally surgical intervention on a part—is also necessary for constitution and, hence, to turn it into a full-blown definition of constitution, in the same way the interventionist theory of causation stipulates that the existence of a change in a putative effect under an ideal intervention on a putative cause is not only sufficient but also necessary for causation. While an in-depth discussion of this proposal is beyond the scope of this paper, we conclude that the general project of using interventions to analyze constitution still holds promise.

The paper is structured as follows. Section 2 introduces the features of constitution and the metaphysical principles we need for our proposal. Section 3 motivates and defines horizontal surgicality, whose use for analyzing constitution is then discussed in Section 4.

2 Preliminaries

Constitution is a dependence relation characterized by a number of features and principles many authors explicitly or implicitly endorse. In what follows, we render transparent the features and principles that will be relevant for our subsequent argument.

First, constitution relates upper-level phenomena and their lower-level constituents, both of which are types of behavior exhibited by *specific entities* on upper and lower levels, respectively.[5] We represent such behaviors by *specific variables* as introduced by Spohn (2006). Specific variables represent behaviors of specific entities

[4] A cause I of X is a fat-handed intervention on X w.r.t. Y when it violates condition (I3) of Woodward's (2003, p. 98) definition of an ideal intervention, such that I causes Y *along two (or more) different paths* (cf. Scheines, 2005, pp. 931-32). The contrast class of fat-handed interventions is the class of surgical interventions. Notice that the distinction between surgical and fat-handed interventions is orthogonal to that between structural and parametric interventions (Eberhardt and Scheines, 2007, pp. 986-87), namely between interventions that, respectively, do and do not satisfy condition (I2).

[5] It is a common (often implicit) background assumption in the mechanistic literature that the relata of constitution are not gerrymandered behaviors, even though this assumption is not underwritten by an explicit criterion for gerrymanderedness (Franklin-Hall, 2016, §5). We, too, assume that all analyzed vari-

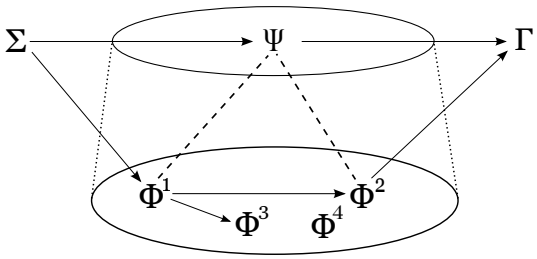

Fig. 1 Ψ is a phenomenon with two constituents, Φ^1 and Φ^2, and two non-constituting parts, Φ^3 and Φ^4. Dashed lines represent constitution, directed edges symbolize causation, and the dotted lines stand for spatiotemporal overlap.

embedded in specific (mechanistic) contexts (e.g., behaviors of specific transistors contained in specific amplifiers, or of specific neurons contained in specific brains, etc.). More concretely, we use Ψ to stand for the behavior of interest of the upper-level entity s. Since each lower-level entity can perform many behaviors, we need a notation that allows us to express this entity- and behavior-relativity at once. Hence, Φ^a_m, Φ^b_n, etc. shall denote the behaviors of type m, n, etc. of lower-level entities x_a, x_b, etc. (for the reader to remember: $\Phi^{entity}_{behavior}$). In general, $\Phi^k_i = j$ shall indicate the fact that some entity x_k, where k ranges over the various entities contained in s, exhibits some type of behavior Φ_i, where i ranges over possible activities, in some particular token way j, where j ranges over the possible token values of Φ_i. To denote behaviors of entities s_k outside of a mechanistic system we use other Greek letters, A^k_i, B^k_i, Γ^k_i, etc. For convenience, we moreover define the *constituting set* $\mathbf{\Phi}$ of a phenomenon Ψ to be the set of all and only the constituents responsible for Ψ (relative to their mechanistic—spatial, temporal, and causal—context; see Craver, 2007, p. 153).

To illustrate, Figure 1 depicts a mechanistic system (where we assume, for simplicity, that every entity/part exhibits only one behavior): the phenomenon Ψ in the upper-level ellipse contains four behaviors in the lower-level ellipse; only two of them, *viz.* Φ^1 and Φ^2, are constituents of Ψ. That abstract structure can, for instance, be interpreted in terms of the mechanism underlying the amplification phenomenon exhibited by a two-stage serial amplifier. In a two-stage amplifier, a voltage difference is applied to a first transistor, which amplifies the signal and feeds it to a second transistor, which also amplifies the signal and outputs it to some other device, say, a loudspeaker. In principle, the overall gain is the sum of the individual gains of the two transistors. Moreover, each transistor produces heat, which may require heat sinks in order to prevent overheating. The signal subject to amplification is a complex wave made up of many different frequencies and amplitudes. Depending on the kind and position of the circuits' elements, the amplifier may introduce an audible distortion of the signal's waveform. Against that background, we may interpret Σ in Figure 1 as the voltage difference applied to the amplifier, Ψ as the overall gain of the amplifier, and Γ as the audible distortion; Φ_1 and Φ_2 are the individual gains of the first and

ables represent non-gerrymandered behaviors and that it is pre-theoretically clear what gerrymanderedness amounts to.

second transistor. The amplifier also features behaviors that are constitutively irrelevant to its gain, *viz.* the behavior Φ^3 of the heat sink of the first transistor, and the behavior Φ^4 of a malfunctioning heat sink, which is physically (and thus causally) detached from the second transistor.

Second, as is common for the mechanistic literature, we will henceforth assume clarity on the spatiotemporal parthood relations obtaining among analyzed behaviors; that is, in particular, for any two behavior instances $\Phi_i^h = i$ and $\Phi_i^k = j$ relative to some phenomenon of interest $\Psi = l$, it is clear whether the states i and j stand in a relation of proper parthood, meaning that one of i and j occupies a smaller spacetime region than the other and is contained in the spacetime region occupied by the other. In the vein of Eronen (2013, p. 1047), we define $\Phi_i^k = j$ to be a *direct* proper part of $\Psi = l$ iff j is a proper part of l and there does not exist another behavior $\Phi_i^h = i$ contained in $\Psi = l$ such that j is a proper part of i. This gives us a hierarchy of direct proper parthood, which can be used to locally distinguish *spatiotemporal levels* of Ψ: Ψ is on the top level; the variables representing direct proper parts of Ψ are on the next lower level; then come the variables representing direct proper parts of the direct proper parts of Ψ, and so on.[6] Against that background, the relation between a phenomenon Ψ and its constituents in $\mathbf{\Phi}$ can be more specifically characterized by the following *Parthood* principle: the instances of the elements of $\mathbf{\Phi}$ are spatiotemporal parts of instances of Ψ, that is, the spatiotemporal region occupied by an instance of the phenomenon contains the spatiotemporal regions occupied by the instances of the constituents (Leuridan, 2012, p. 410). Those elements of $\mathbf{\Phi}$ that represent direct parts of Ψ constitute Ψ on the first lower level, those representing direct parts of the first-level constituents are second-level constituents, those representing direct parts of the second-level constituents third-level constituents, and so forth.

Third, the relation between a phenomenon and its constituents is to be analyzed in terms of *non-reductive supervenience* (Glennan, 1996, pp. 61-2; Eronen, 2011, ch. 11). More specifically, a phenomenon Ψ supervenes on $\mathbf{\Phi}$, meaning that every change in Ψ is necessarily accompanied by a change in at least one $\Phi_i^k \in \mathbf{\Phi}$ (relative to a given mechanistic context), such that, due to multiple realizability, the phenomenon is not reducible—in particular not identical—to its constituents. Moreover, changes in Ψ and in its supervenience base occur simultaneously, and we assume that phenomena can, in principle, be constitutionally explained on any of their lower levels. That, in turn, presupposes that there exist constituents of a phenomenon on every lower level, meaning there are no gaps in constitutional hierarchies. We combine these features of constitution into a principle we call *Universal Constitution*:

(UC) Every (change in a) state of any (non-fundamental) phenomenon Ψ is necessarily and simultaneously realized by (a change in) the state of at least one constituent Φ_i^k of Ψ on every lower level, such that $\Psi \neq \Phi_i^k$.

It follows from UC that free-floating spatiotemporal phenomena, which are not constitutively anchored or grounded, are impossible (with the exception of phenomena on a bottom, fundamental level, if such a level exists).

[6] Note that this spatiotemporal notion of a level is merely instrumental for our ensuing argument; it is not intended as a contribution to the ongoing debate on levels in the mechanistic literature nor, in particular, as an alternative to Craver's (2007, p. 189) or Bechtel's (2008, p. 146) notions of a level.

Finally, we assume *Simultaneity*:

(SIM) Simultaneous changes cannot be causally related.

As has been argued by Craver and Bechtel (2007), SIM, in combination with the fact that changes in phenomena and constituents are simultaneous, which follows from UC, entails that constitution is a non-causal form of dependence. Causation relates mereologically independent entities (cf. Woodward and Hausman, 1999, p. 523), such that causes temporally precede their effects. In contrast, constitution holds among wholes and their parts, that is, among spatiotemporally overlapping entities, such that changes in phenomena necessarily co-occur with changes in their constituents, and thus cannot—*pace* Leuridan (2012)—be causally related to them.

SIM and UC entail *Systematic Fat-Handedness*: phenomena and their constituents can only be manipulated with a fat hand, that is, via common causes.[7] To see this, suppose that Σ is a cause of a phenomenon Ψ. By supervenience, all changes Σ induces on Ψ are necessarily associated with changes in at least one constituent $\Phi_i^k \in \mathbf{\Phi}$, meaning that Σ not only makes a difference to Ψ but also to Φ_i^k. It follows from difference-making theories of causation, of which Woodward's (2003) interventionism is the most prominent exemplar, that Σ not only causes Ψ but also Φ_i^k, meaning that there is at least one causal path from Σ to Ψ and Φ_i^k. For mere logical reasons, this can be structurally realized in one of two ways: either Σ causes Ψ and Φ_i^k along *one* causal path, e.g. $\Sigma \longrightarrow \Psi \longrightarrow \Phi_i^k$, or along *more than one* path, $\Psi \longleftarrow \Sigma \longrightarrow \Phi_i^k$. The former option is excluded by the fact that the changes in Ψ and Φ_i^k are simultaneous and their relationship, hence, is non-causal. In light of the non-identity of phenomena and their parts and the standard definition of (directed) causal paths in terms of ordered n-tuples of variables (Spirtes, Glymour, and Scheines, 2000, pp. 8-9), it follows that Σ causes Ψ and Φ_i^k along two different paths, *viz.* $\langle \Sigma, \Psi \rangle$ and $\langle \Sigma, \Phi_i^k \rangle$ with $\Psi \neq \Phi_i^k$, meaning that Σ is a common cause of Ψ and Φ_i^k. This argument, of course, generalizes: every cause of a phenomenon is also a cause of some constituent or other of that phenomenon on a different causal path (Romero 2015; Baumgartner and Gebharter, 2016).

3 Horizontal surgicality

In the light of Systematic Fat-Handedness, the ultimate deficiency of MM is easily pinpointed: MM definitionally ties the notion of constitution to the possibility of ideal interventions that target *one* level of a mechanistic system and thereby change the other level, where in fact it is only possible to induce changes on upper and lower levels of a mechanism by fat-handedly targeting *both* levels on different causal paths. Contrapositively put, whenever ideal interventions that target a first variable and induce changes in a second one are possible, these variables are not linked in terms of constitution but in terms of causation—as is duly entailed by the interventionist theory of causation (Woodward, 2003). The obvious conclusion to draw is that constitution must not be analyzed in terms of ideal interventions. In what follows, we develop a weakened notion of an intervention based on which we will first provide a

[7] A similar point is made by Eronen and Brooks (2014), without the terminology of fat-handedness.

sufficient condition for constitution and then, in Section 4, reconsider the possibility of a full-blown interventionist definition of constitution.

On the one hand, Systematic Fat-Handedness suggests that in order to be able to intervene on mechanistic systems, some common causes must pass as interventions. On the other hand, as fat-handed interventions generate confounded data and, thus, come with greatly diminished inferential leverage (cf. e.g. Scheines, 2005), there are good methodological reasons not to admit *all* common causes into the category of (potential) interventions. In fact, methodological prudence calls for maximal restrictiveness in assigning the status of an intervention to common causes.

Prompted by problems of the original version of interventionism with macro-to-micro causation, Woodward (2015) has recently weakened his original notion of an ideal intervention by introducing exemption clauses for supervenience relations. While he required in (2003, p. 98) that an intervention targets exactly one variable, he now (2015, pp. 333-34) allows for multiple targets, provided that these targets are related in terms of supervenience. Woodward contends "that an intervention on a macro variable Ψ also should be treated as automatically changing (indeed as also an intervention on) the supervenience base SB(Ψ) of Ψ" (2015, p. 333, adjusted to our symbolism). Thus, if phenomena and their constituents are related by supervenience, as is standardly assumed in mechanistic theorizing (and as we assume for the purposes of this paper, cf. Section 2), interventions that target both phenomena and constituents on different causal paths pass as interventions, notwithstanding the fact that they are non-surgical but fat-handed. For brevity, we will speak of *permissibly* fat-handed interventions. The notion of a permissibly fat-handed intervention imposes a far-reaching restriction on common causes to pass as interventions. Common causes whose parallel effects are not related by some residual dependence—as is the case for most common causes—do not count as interventions. Only common causes with parallel effects that are related in terms of supervenience can figure as interventions.

As phenomena and their constituents stand in a supervenience relation, their common causes pass as permissibly fat-handed interventions. It follows that against the background of Woodward's weakened notion of an intervention mutual manipulability of macro and micro levels is no longer ruled out for mere conceptual reasons.[8] Let us thus ask whether it is possible to get an adequate account of constitution by simply replacing MM's ideal interventions with *permissibly* fat-handed interventions, in the following vein:

(pFAT) Φ_i^k constitutes Ψ iff

(i) the instances of Φ_i^k are spatiotemporal parts of instances of Ψ

(ii) there exists a possible *permissibly* fat-handed intervention $I_{\Phi_i^k} = i$ on Φ_i^k w.r.t. Ψ that causes $\Phi_i^k = j$, for some j, and simultaneously changes Ψ;

(iii) there exists a possible *permissibly* fat-handed intervention $I_\Psi = j$ on Ψ w.r.t. Φ_i^k that causes $\Psi = l$, for some l, and simultaneously changes Φ_i^k.

[8] Notice, however, that Woodward (2015) does not modify the notion of an intervention for the purpose of testing for constitution but rather for the purpose of testing for causation in variable sets including supervenience relations.

Unfortunately, the question must be answered in the negative, as pFAT is question-begging. Whether some common cause I of Ψ and Φ_i^k counts as a permissibly fat-handed intervention depends on whether Φ_i^k belongs to the supervenience base of Ψ, which, in turn, depends on whether Φ_i^k is a constituent of Ψ. That I is a permissibly fat-handed intervention on Ψ and Φ_i^k presupposes—rather than establishes—that Φ_i^k is a constituent of Ψ. This problem, of course, generalizes to any other attempt to define constitution on the basis of permissibly fat-handed interventions. The notion of a permissibly fat-handed intervention presupposes clarity on the supervenience relations among the variables in a scrutinized system. By contrast, a theory of constitution aims to provide such clarity by identifying the micro constituents forming the supervenience base of macro variables.

In light of the uninformativeness of pFAT, it might be suggested that a qualification of the notion of a fat-handed intervention is redundant because it is possible to account for constitution on the basis of fat-handed interventions *tout court*. However, it is easily seen that a theory that recognizes *all* common causes as potential interventions and disregards the italicized qualifications in pFAT—let us call this theory FAT—is a non-starter. The mere existence of common causes of upper and lower levels of a mechanism is insufficient to distinguish between constituting and non-constituting parts. Just like constituting parts, non-constituting parts, too, may share common causes with phenomena.

For example, reconsider the mechanism in Figure 1, in which the lower-level variable Φ^3 represents the behavior of the (functioning) heat sink of the first transistor. Φ^3 is an effect of the constituent Φ^1, but does not itself constitute the phenomenon Ψ. The heat sink's behavior is, using Craver's (2007, p. 143) jargon, a *sterile effect* of the mechanism underlying the amplification, that is, a downstream effect of a constituent that plays no role in the bottom-level realization of the phenomenon. As Φ^3's direct causal parent Φ^1 and Ψ are systematically coupled via common causes in the structure of Figure 1, Φ^3 and Ψ likewise share a multitude of common causes, which, in turn, yields that many changes in the former will be associated with changes in the latter. For instance, inducing a voltage difference causes heat to be transferred to the sink and it causes the amplification, and, conversely, stopping the voltage difference terminates both the heating and the amplification. Nonetheless, of course, the heat sink's behavior is not a constituent of the mechanism for amplification. The reason is that the influence of the common causes of amplification and the heat sink's behavior is mediated via another spatiotemporal part of the amplifier, namely the first transistor's behavior, which *is* a constituent of amplification. This example shows that only *simultaneous* changes in parts and phenomena are revealing of constitution.

Moreover, consider the lower-level variable Φ^4, which represents the behavior of the malfunctioning heat sink, disconnected from the mechanism. A hammer can smash both the heat sink and, say, the first transistor and, as a result of this, change the amplification phenomenon. But clearly, the existence of a common cause (*viz.* a fat-handed intervention) of the disconnected heat sink's behavior and the amplification does not entail that the former is a constituent of the latter. The reason is that the cause is ham-fisted on the lower level, as it not only crashes the disconnected part but also the first transistor, whose behavior Φ^1 *is* a constituent of amplification. That is, an intervention that is revealing of constitution must not directly target multiple

parts of a scrutinized phenomenon *on the same level*. Still, constitutionally revealing fat-handed interventions on mechanistic systems need not be surgical with respect to all parts of a phenomenon Ψ *on all levels*. As Parthood allows for mereological hierarchies, parts of Ψ can be parts of other parts of Ψ on other levels. Since causes inducing changes on spatiotemporally overlapping processes are common causes of these processes, and thus non-surgical, every intervention on Ψ is a common cause of Ψ and of at least one of its parts Φ_i^k on a first lower level, and of at least one of Φ_i^k's parts on a second lower level, and so on. Moreover, by UC, some of these changes will be necessary, such that any intervention on Ψ necessarily targets, on different causal paths, a multitude of Ψ's parts on all lower levels. A horizontally surgical intervention $I_{\Phi_i^k}$ on Φ_i^k with respect to Ψ must thus be defined in such a way that $I_{\Phi_i^k}$ is allowed to cause changes in multiple parts of Ψ, provided that $I_{\Phi_i^k}$ does not directly target *more than one* part of Ψ on every lower level.

In sum, fat-handed interventions, without further qualifications, are insufficient to infer to constitution. Baumgartner and Casini (2017) have recently proposed an abductive theory of constitution based on fat-handed interventions, which—without making use of Simultaneity—aims to compensate for the shortcomings of FAT. On the one hand, they find that by gradually expanding analyzed variable sets it becomes possible to identify sterile effects and disconnected parts on interventionist grounds but, on the other, they show that unrestricted fat-handed interventions never ground a conclusive inference to constitution. In this paper, we take a different approach: we impose a further restriction on fat-handed interventions—*viz*. horizontal surgicality—to turn them into constitutionally revealing tools. More specifically, based on the above considerations, we propose the following definition of a horizontally surgical intervention:

(H) $I_{\Phi_i^k}$ is a horizontally surgical intervention variable on a part Φ_i^k of Ψ w.r.t. Ψ iff:

- (i) $I_{\Phi_i^k}$ is a cause of Φ_i^k;
- (ii) if $I_{\Phi_i^k}$ causes changes in both Φ_i^k and Ψ, these changes occur simultaneously;
- (iii) $I_{\Phi_i^k}$ is a direct cause of at most one behavior on every level lower than Ψ's.

We shall then say that $I_{\Phi_i^k} = i$ is a horizontally surgical intervention on a proper part Φ_i^k of Ψ w.r.t. Ψ iff $I_{\Phi_i^k} = i$ fixes Φ_i^k to some value j, such that it simultaneously changes the value of Ψ, and directly changes at most one behavior on every level lower than Ψ's.

The next section will show that fat-handed interventions on mechanistic systems that are horizontally surgical in this sense are indeed revealing of constitution.

4 Mechanistic constitution

We claim that for a part Φ_i^k of a phenomenon Ψ to constitute Ψ (i.e. for Φ_i^k to belong to the set $\mathbf{\Phi}$ of Ψ's constituents), it suffices that the following condition be satisfied:

(SUF) There exists a (possible) horizontally surgical intervention $I_{\Phi_i^k} = i$ on Φ_i^k w.r.t. Ψ that causes changes in both Φ_i^k and Ψ.

For the ensuing proof that SUF is indeed sufficient for constitution we need, in addition to the principles introduced in Section 2, a principle of *Transitivity* about the co-occurrence of events (at non-relativistic distances):

(TR) If x co-occurs with y, and y co-occurs with z, then x co-occurs with z.

The proof can be intuitively motivated as follows. Since every (higher-level) phenomenon is constituted by some behavior on every lower level, any intervention that changes the phenomenon simultaneously changes one of the constituents on every lower level. Now, if an intervention on the phenomenon leads to simultaneous changes in the phenomenon and a *non*-constituent on a specific level, this intervention necessarily also changes at least one constituent on that level, in addition to the non-constituent. Horizontally surgical interventions are defined such that they change maximally one behavior on each level simultaneously with the phenomenon. Hence, any intervention on a phenomenon that changes the phenomenon and, at the same time, a non-constituent cannot be horizontally surgical. The proof is thus a *reductio ad absurdum* of the negation of the conditional "If there exists a horizontally surgical intervention on Φ_i^k w.r.t. Ψ causing changes in both Φ_i^k and Ψ, then Φ_i^k is a constituent of Ψ". That is, it deduces a contradiction from assuming the truth of the conditional's antecedent (premise 1 below) and the falsehood of its consequent (premise 2 below).

(1)	$I_{\Phi_i^k} = i$ is a horizontally surgical intervention on Φ_i^k w.r.t. Ψ that causes changes in both Φ_i^k and Ψ.	ASM
(2)	$\Phi_i^k \notin \mathbf{\Phi}$.	ASM
(3)	The changes induced on Ψ and Φ_i^k by $I_{\Phi_i^k} = i$ are simultaneous.	1, H
(4)	The change induced on Ψ by $I_{\Phi_i^k} = i$ is realized by a simultaneous change in at least one part Φ_j^h of Ψ, such that $\Phi_j^h \in \mathbf{\Phi}$, that $\Phi_i^k \neq \Phi_j^h$, and that Φ_i^k and Φ_j^h are on the same level.	2, UC
(5)	The changes in Φ_i^k and Φ_j^h are simultaneous.	3, 4, TR
(6)	The changes induced on Φ_i^k and Φ_j^h by $I_{\Phi_i^k} = i$ are not mutually causally related, i.e. they are the result of a common cause structure, *viz.* $\Phi_j^h \longleftarrow I_{\Phi_i^k} \longrightarrow \Phi_i^k$.	5, SIM
(7)	$I_{\Phi_i^k} = i$ directly causes two behaviors on the same level.	4,6
(8)	$\bot$	1, 7

To explain how the contradiction arises with the aid of an illustration, assume—for *reductio*—that (1) an intervention $I_{\Phi_i^k} = i$, say a needle pinching the hippocampus of a rat navigating a water maze, changes the hippocampus' behavior of generating spatial maps in the rat's brain, Ψ, as well as the oxygen concentration in some blood vessel of the hippocampus, Φ_i^k, such that the intervention on Φ_i^k w.r.t. Ψ is horizontally surgical; and yet that (2) Φ_i^k is *not* a constituent of Ψ. By H, it follows that (3) the changes in Φ_i^k and Ψ are simultaneous. As Φ_i^k does not constitute Ψ, it follows, by UC, that (4) some constituent Φ_j^h must simultaneously realize the change

in Ψ on the same level as Φ_i^k (because phenomena must be constitutively realized on all lower levels).[9] In our example, Φ_j^h can be interpreted as the firing of some neurons in the hippocampus. Since also the changes in Ψ and Φ_j^h, and not only the changes in Φ_i^k and Ψ, are simultaneous, it follows, by TR, that (5) the changes in Φ_j^h and Φ_i^k are simultaneous, too. By SIM, it follows that (6) the changes in Φ_i^k and Φ_j^h are not mutually causally related, meaning they are brought about by a common cause structure $\Phi_j^h \longleftarrow I_{\Phi_i^k} \longrightarrow \Phi_i^k$, and not via a directed path $I_{\Phi_i^k} \longrightarrow \Phi_i^k \longrightarrow \Phi_j^h$ or $I_{\Phi_i^k} \longrightarrow \Phi_j^h \longrightarrow \Phi_i^k$. That is, neural activity and oxygen concentration are changed by the intervention on separate paths. That, in turn, entails that (7) $I_{\Phi_i^k} = i$ directly causes two behaviors on the same level, *viz.* Φ_i^k and Φ_j^h, which, as stated in (8), contradicts the initial assumption that $I_{\Phi_i^k} = i$ is horizontally surgical due to a violation of condition H(iii).

To resolve the contradiction in (8), at least one of the premises has to go, meaning that we must reject one of the principles UC, TR, and SIM, or one of the assumptions (1) and (2). We have already defended UC and SIM. TR is widely accepted in the macroscopic—yet not cosmic—domain of investigation of the special sciences, that is, in the domain where mechanistic explanations are appropriate. Moreover, although the blood vessels' behaviors are spatiotemporally contained in the hippocampus' map generation, the latter is commonly taken to be constituted by neural activity and not by the oxygen concentration in the blood vessels surrounding the neurons. The oxygen concentration is, if anything, a sterile effect of neural activity (see Craver, 2007, pp. 151-52, and references therein). Hence, (2) likewise cannot be rejected. The only viable option, therefore, is to deny (1). Since oxygen concentration is not a constituent, the change in oxygenation cannot possibly simultaneously co-vary with the phenomenon unless neural activity also simultaneously co-varies with it, meaning that the intervention is not horizontally surgical. The upshot is that, if Φ_i^k is not a constituent of Ψ, then the intervention $I_{\Phi_i^k} = i$ on Φ_i^k w.r.t. Ψ cannot be horizontally surgical. Since nothing hinges on $I_{\Phi_i^k} = i$ being our candidate intervention, the argument generalizes: there does not exist a non-constituent for which there exists a horizontally surgical intervention; or contrapositively, all parts for which horizontally surgical interventions exist are constituents.[10]

It is tempting to build a full-blown interventionist definition of constitution on the fundament of SUF by stipulating that, in combination with Parthood, SUF is not only sufficient but also necessary for constitution:

(hFAT) Φ_i^k constitutes Ψ iff

(i) the instances of Φ_i^k are spatiotemporal parts of instances of Ψ;

(ii) there exists a (possible) horizontally surgical intervention $I_{\Phi_i^k} = i$ on Φ_i^k w.r.t. Ψ that causes changes in both Φ_i^k and Ψ.

[9] The behaviors represented by Φ_i^k and Φ_j^h differ either because they involve different entities or different activities, that is, either because of $i \neq j$ or because of $h \neq k$.

[10] Relative to a different interpretation of the variables in the reductio, the resulting contradiction can, of course, also be resolved by rejecting (2) and upholding (1). If Φ_i^k is interpreted to stand for neural activity and not for oxygenation, Φ_i^k turns out to be a constituent of Ψ, meaning that assumption (2) has to go.

hFAT imposes a Parthood condition, like MM, but replaces top-down and bottom-up manipulations with horizontally surgical interventions. Moreover, hFAT is built in analogy to Woodward's (2003) interventionist definition of causation, which also upgrades a provably sufficient condition for causation, *viz.* the existence of a difference-making scenario induced by an ideal intervention, to a full-blown definition by stipulating that this condition is not only sufficient but also necessary for causation. However, this feature of interventionism has given rise to well-known objections, as it is questionable whether ideal interventions indeed exist for *every* causal dependence. For instance, it is physically impossible to ideally intervene on the moon w.r.t. the earth's tides (Woodward, 2003, p. 129) or on Big Bang w.r.t. the latter's downstream effects (Maudlin, 2002, p. 149). Discussing the adequacy of the interventionist account of causation is beyond the scope of this paper. For our current purposes it suffices to stress that hFAT is subject to the same vulnerabilities, insofar as the existence of horizontally surgical interventions depends, minimally, on the existence of ideal interventions on any putative constituent w.r.t. spatiotemporally non-overlapping parts of the phenomenon under scrutiny. But if the (possible) existence of such interventions is questionable in causal contexts, it is, of course, likewise questionable in constitutional contexts.

Still, it is important to emphasize that hFAT is not subject to MM's conceptual flaws. While MM defines constitution in terms of interventions that cannot possibly exist, and thus entails the paradoxical conclusion that there cannot possibly exist instances of constitution, hFAT has no such implications. The adequacy of hFAT, just as the adequacy of the interventionist definition of causation, is a matter of *scope*, in the sense that the definition may be subject to counterexamples, but it does not have any paradoxical ramifications to the effect that the existence of instances of the definiens is excluded on a priori grounds. This, we submit, is a welcome step forward for anyone, who wants to analyze constitution in terms of (some form of) interventions.

5 Conclusion

Recent literature has shown that definitions of constitution based on the notion of ideal interventions are flawed: if constitution is tied to the possible existence of ideal interventions, either constitution reduces to causation, contrary to the widespread view that constitution is different from causation, or cases of constitution cannot possibly exist, because the ideal interventions on phenomena, which non-reductively supervene on mechanisms, are unrealizable in principle. While several authors have taken these criticisms as motivating non-interventionist analyses of constitution, we further explored the chances of reviving the interventionist project. To this end, we introduced a novel notion of intervention—*viz.* the notion of a *horizontally surgical* intervention—that, we proved, provides a *sufficient condition* for constitution. Furthermore, we argued that if that condition is also taken to be necessary for constitution, a full-blown interventionist *definition* of constitution results, which is not subject to the conceptual flaws of definitions in terms of ideal interventions.

Acknowledgements We thank the participants to the Biological Interest Group of the University of Geneva, 21 February 2017, and the audiences of BSPS, Edinburgh, 13 July 2017, and SMS, New York, 5 October 2017. This research was generously supported by the Swiss National Science Foundation, grants no. 100017_169810 and 100012E_160866/1 for LC and grant no. PP00P1_144736/1 for MB. MB is moreover indebted to the Toppforsk-programme of the Bergen Research Foundation and the University of Bergen, grant no. 811886. BK was supported by the DFG-Graduiertenkolleg "Situated Cognition", GRK-2185/1.

References

Baumgartner, M. and L. Casini (2017). An abductive theory of constitution. *Philosophy of Science 84*(2), 214–33.

Baumgartner, M. and A. Gebharter (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science 67*(3), 731–56.

Bechtel, W. (2008). *Mental Mechanisms*. London: Routledge.

Couch, M. B. (2011). Mechanisms and constitutive relevance. *Synthese 183*, 375–88.

Craver, C. and W. Bechtel (2007). Top-down causation without top-down causes. *Biology and Philosophy 22*, 547–63.

Craver, C. and J. Tabery (2017). Mechanisms in science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 ed.). Metaphysics Research Lab, Stanford University.

Craver, C. F. (2007). *Explaining the Brain*. Oxford: Oxford University Press.

Eberhardt, F. and R. Scheines (2007). Interventions and causal inference. *Philosophy of Science 74*, 981–995.

Eronen, M. I. (2011). *Reduction in Philosophy of Mind: A Pluralistic Account*. Frankfurt am Main: Ontos.

Eronen, M. I. (2013). No levels, no problems: Downward causation in neuroscience. *Philosophy of Science 80*(5), 1042–52.

Eronen, M. I. and D. S. Brooks (2014). Interventionism and supervenience: A new problem and provisional solution. *International Studies in the Philosophy of Science 22*(2), 185–202.

Franklin-Hall, L. R. (2016). New mechanistic explanation and the need for explanatory constraints. In K. Aizawa and C. Gillett (Eds.), *Scientific Composition and Metaphysical Ground: New Directions in the Philosophy of Science*, Chapter 2. London: Springer.

Gebharter, A. (2017). Uncovering constitutive relevance relations in mechanisms. *Philosophical Studies 174*, 2645–66.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis 44*, 49–71.

Harbecke, J. (2010). Mechanistic constitution in neurobiological explanations. *International Studies in the Philosophy of Science 24*, 267–85.

Harbecke, J. (2015). The regularity theory of mechanistic constitution and a methodology for constitutive inference. *Studies in History and Philosophy of Biological and Biomedical Sciences 54*, 10–19.

Irvine, E. (2013). *Consciousness as a Scientific Concept*. Dordrecht: Springer.

Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology and Philosophy 27*, 545–70.

Krickel, B. (2018). Saving the mutual manipulability account of constitutive relevance. *Studies in History and Philosophy of Science Part A 68*, 58–67.

Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *The British Journal for the Philosophy of Science 63*, 399–427.

Machamer, P., L. Darden, and C. Craver (2000). Thinking about mechanisms. *Philosophy of Science 67*, 1–25.

Maudlin, T. (2002). *Quantum Non-locality and Relativity*. Oxford: Blackwell.

Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese 192*(11), 3731–55.

Scheines, R. (2005). The similarity of causal inference in experimental and non-experimental studies. *Philosophy of Science 72*(5), 927–40.

Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search* (second ed.). Cambridge MA: MIT Press.

Spohn, W. (2006). Causation: An alternative. *The British Journal for the Philosophy of Science 57*, 93–119.

van Eck, D. and H. L. de Jong (2016). Mechanistic explanation, cognitive systems demarcation, and extended cognition. *Studies in History and Philosophy of Science Part A 59*, 11–21.

Woodward, J. (2003). *Making Things Happen. A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research 91*, 303–47.

Woodward, J. and D. Hausman (1999). Independence, invariance and the causal Markov condition. *The British Journal for the Philosophy of Science 50*(4), 521–83.

Zednik, C. (2015). Heuristics, descriptions, and the scope of mechanistic explanation. In C. Malaterre and P.-A. Braillard (Eds.), *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*, pp. 295–318. Dordrecht: Springer.